



TITLE:

High-quality Knowledge Acquisition of
Predicate-argument Structures for Syntactic
and Semantic Analysis(Abstract_要旨)

AUTHOR(S):

Jin, Gongye

CITATION:

Jin, Gongye. High-quality Knowledge Acquisition of Predicate-argument Structures for Syntactic and Semantic Analysis. 京都大学, 2016, 博士(情報学)

ISSUE DATE:

2016-03-23

URL:

<https://doi.org/10.14989/doctor.k19850>

RIGHT:

(続紙 1)

京都大学	博士（情報学）	氏名	金 弓治
論文題目	High-quality Knowledge Acquisition of Predicate-argument Structures for Syntactic and Semantic Analysis (構文・意味解析のための高品質な述語項構造知識の獲得)		
(論文内容の要旨)			
<p>This dissertation describes a series of approaches for acquiring high-quality knowledge of predicate-argument structures, which is indispensable for text understanding by computers. The dissertation also proposes methods of syntactic and semantic analysis based on the acquired knowledge of predicate-argument structures. The dissertation is composed of seven chapters, out of which five chapters are devoted to describe the proposed approaches.</p> <p>Chapter 1 introduces knowledge acquisition approaches and fundamental analysis methods that are used for knowledge acquisition. Knowledge acquisition is tightly dependent on fundamental analyses, such as part-of-speech tagging and syntactic parsing. Dependency parsing, in particular, has been widely employed for knowledge acquisition. Although the accuracy of state-of-the-art dependency parsers for some languages, such as English or Japanese, is over 90%, it is still not high enough to acquire precise knowledge. Furthermore, for some difficult-to-analyze languages, such as Chinese and Arabic, the performance becomes much worse. When we apply this kind of parsing models to real-world texts, such as Web pages, more performance drops are inevitable. This observation leads to the motivation of the approaches proposed in later chapters. An overview of the proposed approaches is also given.</p> <p>Chapter 2 describes a language-independent method for selecting high-quality dependencies from automatic parses to deal with errors in dependency parsing. This method makes use of a supervised classifier based on various linguistic features extracted from dependency trees. The experimental results on seven languages including Chinese showed that the proposed method can effectively select high-quality dependencies from automatic parses.</p> <p>Chapter 3 introduces a language-independent framework for automatically acquiring high-quality knowledge of predicate-argument structures, which includes a set of predicate-argument structures and case frames. Case frames represent relations between a predicate and its arguments, and the product of clustering the uses of predicate-argument structures. The proposed framework enables the automatic compilation of case frames in multiple languages. In the experiments, English and Chinese case frames were compiled, and their quality was proved to be high by manual evaluation.</p> <p>Chapter 4 describes a method for applying the knowledge of predicate-argument</p>			

structures to dependency parsing. So far, semi-supervised approaches to dependency parsing have suffered from noises in additional knowledge. To overcome this problem, the proposed method uses the high-quality knowledge of predicate-argument structures as features of a graph-based dependency parser. Experimental results on Chinese dependency parsing showed that the performance of dependency parsing was significantly improved by the high-quality knowledge, especially by the knowledge extracted from a huge Web corpus.

Chapter 5 presents a method for integrating the knowledge of predicate-argument structures into semantic role labeling, which is a kind of semantic analysis. For semantic role labeling, semi-supervised approaches have not been successful due to the lack of high-quality knowledge that is effective for semantic role labeling. To solve this problem, the proposed method incorporates the high-quality knowledge of predicate-argument structures into each stage of the flow of semantic role labeling as features of machine learning. Experimental results on Chinese semantic role labeling showed that higher-quality knowledge has a larger effect on improving the performance of semantic role labeling.

Chapter 6 describes a method for selecting high-quality semantic roles from the results of semantic role labeling. Since the process of semantic role labeling consists of a sequence of several analysis stages, the outputs of semantic role labeling inevitably contain many errors propagated from the lower-level stages. To avoid these errors, selecting high-quality semantic roles from the output of semantic role labeling is necessary for semantic knowledge acquisition based on semantic roles. By the experiments on Chinese, the proposed method was proved to effectively select high-quality semantic roles. These high-quality semantic roles can be used to compile deep-level case frames based on semantic roles.

Chapter 7 concludes the dissertation and describes future work.

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し

審査結果の要旨は日本語500～2,000字程度で作成すること。

(続紙 2)

(論文審査の結果の要旨)

本論文は、計算機によるテキスト理解を実現する上で不可欠である高品質な述語項構造知識を自動獲得する手法、および自動獲得した述語項構造知識に基づく構文・意味解析手法について研究し、その成果をまとめたものである。述語項構造知識とは、述語項構造の集合や、述語項構造をクラスタリングした格フレームであり、構文解析、機械翻訳などの言語処理システムの高度化、さらにはテキスト理解において重要な知識源である。述語項構造知識の自動獲得は、形態素解析、構文解析などの基礎的な解析技術に基づいているため、自動獲得した述語項構造知識の質は、これらの基礎解析の精度に大きく依存している。特に、中国語のように基礎解析の精度が低い言語では、獲得した述語項構造知識の質が低くなってしまい、言語処理システムに適用しても精度が向上しないという問題があった。本論文では、この問題を解決することを目的に、主に以下に示す3つの成果を達成した。

1. 高品質な述語項構造知識の獲得に向けて、自動解析結果の中で信頼度が高い解析結果を自動的に選択する手法を提案した。中国語を含む7つの言語に対する実験によって、提案手法が言語に依存せず有効であることを示している。さらに、得られた信頼度の高い解析結果から高品質な述語項構造知識を抽出する手法を開発した。この手法もまた言語に依存せず、構文解析器さえ存在する言語であれば高品質な述語項構造知識を自動抽出することが可能となった。

2. 構文解析の精度向上手法の一つとして、人手で作成した正解データに加えて自動構文解析結果を用いるという半教師有り学習手法が提案されている。しかし、自動構文解析結果には誤りが多く含まれているため、構文解析精度が安定して向上しないという問題があった。本論文では、この問題を解決するために、大規模テキストから自動抽出した高品質な述語項構造知識を用いて構文解析を行う手法を提案した。中国語構文解析実験において、有意に解析精度が向上することを示している。

3. 意味解析の一つである意味役割付与の研究において、意味役割付与に有効な質の高い知識が存在しないため、半教師有り学習はこれまでほとんど成功していなかった。本論文では、大規模な生テキストから自動抽出した高品質な述語項構造知識を利用して意味役割付与を行う半教師有り学習手法を考案した。中国語意味役割付与の実験において、有意に意味役割付与精度が向上することを示している。さらに、意味役割付与の結果から信頼度の高い意味役割を自動的に選択する手法の開発に成功した。これによって、意味的な曖昧性の少ない述語項構造知識を構築することが可能になり、構文・意味解析のさらなる精度向上につながると考えられる。

以上のように本論文は、計算機によるテキスト理解の実現において重要となる知識獲得および解析技術の精度向上に貢献しており、学術上・実用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成28年2月24日に実施した論文内容とそれに関連した試問の結果合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。
更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日： 年 月 日以降